DOCUMENT RESUME

ED 422 386 TM 028 948

AUTHOR Lee, Jaekyung
TITLE Comparative Approach to Evaluating Systemic Reform Policies: Applying Objective Measurement and Multilevel Analysis Methods.
PUB DATE 1998-04-15
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Analysis; *Educational Change; Educational Policy; Elementary Secondary Education; *Evaluation Methods; *Item Response Theory; Measurement Techniques; *State Programs; Systems Analysis
IDENTIFIERS *Multilevel Analysis; *Reform Efforts

ABSTRACT

This study explores an alternative approach to educational program and policy evaluation by using two major educational measurement/analysis methods, and illustrates their integrated applications to evaluating state reform policies. Most evaluations have been done one program at a time, but it is desirable to design evaluation research in a way that compares the effectiveness of several programs that have the same objectives but different content or function on the same set of outcome measures. Applying item response theory to policy and practice survey provides an innovative solution to objective measurement of policies and practices. In addition, multilevel analysis methods would not only provide a means for formulating school and state-level regression models simultaneously but also provide more precise estimates of the extent to which state policies affect school practices. An illustrative study of state policy examines the multilevel linkages between state policies and educational outcomes. First, objective measures of state policies are created through application of the Rasch model. Then the multilevel education policy-practice linkages are examined through the application of the hierarchical linking model. As the results illustrate, the idea of comparing two groups of states on their policy outcome measures is similar to the nonequivalent control group design. However, the research design proposed in this paper differs from the nonequivalent control group design in some significant ways: (1) treatment is not a single program, but a set of programs; (2) group exposure is a matter of degree; (3) all of the programs that constitute treatment do not have to occur between pretest and posttest; and (4) subjects examined on pretest and posttest do not have to be the same, but can be sampled independently. The proposed approach should give more flexibility for evaluation design in real-life settings, but at the same time more difficulties for interpretation of evaluation results. Some concerns are reviewed. (Contains three tables, one figure, and eight references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 422 386

Comparative Approach to Evaluating Systemic Reform Policies: Applying Objective Measurement and Multilevel Analysis Methods

Jaekyung Lee, Ph.D.

College of Education and Human Development

University of Maine

Paper presented at the Annual Meeting of the AERA (San Diego, CA, April 15, 1998)

TM028948

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

JaeKyung Lee

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Introduction

Our body of knowledge around policy/program evaluation derives primarily from studies of discrete federal programs aimed at specific student populations such as compensatory, bilingual, and special education. This knowledge/research base for categorical program evaluation has not caught up with the more recent policy context of pervasive state "standards-based" reforms. During the 1980s and early 1990s, the states increased course credit requirements for graduation, raised standards for teacher preparation, mandated tests for teacher certification, developed state curriculum frameworks or guides, and established new statewide student assessments. These comprehensive state policies aimed at broad student populations, so-called systemic school reforms, considered the effects of change on the total system, and thus are distinctive in terms of the scale and nature of program. This change brings unprecedented challenges to many educational researchers who have mostly conducted one-group, program-by-program evaluation. In light of these concerns, this study explores an alternative approach to educational policy/program evaluation by utilizing two major educational measurement/analysis methods, and illustrates their integrated applications to evaluating state reform policies.

Comparative Approach to Policy/Program Evaluation

Most evaluations have been done one program at a time. Study of a single program can show whether participants are better off after the program than they were before. The classic design for evaluation has been the experimental model. The controlled experiment, however, is often impossible in action settings for two major reasons: 1) the program must serve everybody eligible by mandate; 2) program practitioners believe it is their

professional obligation not to deny service. Even non-equivalent control design hardly becomes the solution due to the difficulty in identifying comparable control students (see Slavin et al., 1989). Moreover, multiple programs are often adopted and implemented at the same time so that it is hardly feasible to sort out the effect of a single program. Therefore, it is more realistic and desirable to design evaluation research in a way that compares the effectiveness of several programs that have the same objectives but different content/function on the same set of outcome measures (see Weiss, 1972).

Because educational policies/programs are rarely set up with conscious and orderly variations for the researchers to study, they should devise methods to capitalize on variations that occur naturally. In the following sections, I will introduce methods that have the potential to address such problems and enhance generalizability of results, and the specification of which strategy under which conditions has better effects with different kinds of participants.

Objective Measurement and Multilevel Analysis Methods

Given interstate variation in educational policies and practices, the American states provide an ideal laboratory for comparative policy evaluation research. Yet there has been little research that systematically examined the linkages between state policies and school practices. The decisive inhibiting factor has been the lack of good measures of educational policies and practices. Variation among states or schools in adopting different policies and practices across time has posed key challenges to evaluation researchers.

Applying the item response theory to policy/practice survey provides an innovative solution to objective measurement of policies and practices. For example, the Rasch measurement model not only specifies the adoption of educational policies/practices as a probability rather than a certainty, but also makes it possible to characterize or compare policy-making/implementation units on an interval scale, independently of policies/practices

adopted by those units (see Wright and Stone, 1979; Lee, 1997a). Further, the misfit analysis would allow us to examine not so much the content validity of a survey instrument as an individual survey unit's peculiar policymaking or implementation pattern.

On the other hand, choosing the unit of analysis plagues the researchers when they get to examine the relationships between different levels of variables. Multilevel analysis methods would not only provide a means for formulating school and state-level regression models simultaneously but also provide more precise estimates of the extent to which state policies affect school practices (see Bryk and Raudenbush, 1992; Lee, 1996). Then we can more reliably identify states where a reform has succeeded, and study them more productively.

Assessing the Impact of State Policies on School Practices

Now I demonstrate how we can apply objective measurement and multilevel analysis methods to evaluating systemic reform policies. This example is drawn from my dissertation research (Lee, 1997b) that examines the multilevel linkages between state policies and educational outcomes. First, I create objective measures of state policies through the application of the Rasch model. Those policy measures are constructed from the 1984 Educational Testing Service and 1991 Council of Chief State School Officers state policy surveys that involve standards-based education reforms (i.e., raising standards for student graduation and teacher certification, developing new state curriculum and assessments). On the other hand, the 1990 and 1992 NAEP Trial State Assessment, the large state 8th grade samples, provides a highly reliable set of school practices in mathematics. Information on the frequency of student-centered, higher-order learning activities is extracted from the NAEP TSA teacher questionnaire that was administered to eighth-grade mathematics teachers.

With these measures of educational policies and practices, I examine the multilevel education policy-practice linkages through the application of the hierarchical linear model (HLM). HLM allows us to partition variance in outcome variable into different levels, explain those variance components with their corresponding levels of predictors, and examine cross-level effects, that is, how variables measured at one level affect relations occurring at another. In this case, a school-level regression model is estimated for the schools in each state to predict the association of school organizational characteristics with instructional practices. Simultaneously, a state-level regression model is also estimated for the states to obtain estimates of the impact of state policies on school practices as well as on the relationship between school conditions and practices. From the HLM analysis of cross-level effects, I find that the impact of state policies on school practices depends on the individual schools' capacity and needs for desired instructional change.

Measuring State Policy Activities in Education Reform as an Independent Variable

The state education policies of the 1980s can be categorized into three major policy areas: curriculum/instruction policies, student standards policies, and teacher standards policies (see Table 1). BIGSTEPS, Rasch measurement program, is used to construct objective measures from the responses of 50 states to policy items: the responses to each policy is dichotomized (yes/no). The 1984 test with 26 policies is reconstructed from the 1984-85 survey initiated by ETS, and the 1991 test with 21 policies is based on the 1991-92 survey initiated by the Council of Chief State School Officers. Although the two tests used different instruments, both cover major state-prescribed educational standards. They include the types of policies in effect in the year of the survey (or legislated by that year but due to become effective after that date). There are four common policies with which I can link those two tests (i.e., credit requirements for graduation, basic skills test, professional skills test, and subject specialty test for entry-level certification).

Table 1. Test Instruments: Measuring State Activism in Standards-based Education Reform

1984 Test	1991 Test
	Curriculum Policies (Content-Driven Reform)
	27. Math Curriculum Framework or Guide 28. Curriculum Framework or Guide Relationship to Math Student Assessment 29. Curriculum Framework or Guide Relationship to Math Textbooks
Student and Teacher Policies (Input-oriented Reform)	
Student Standards Policies	Student Standards Policies
Testing 5. Monitoring 6. Remediation 7. Gatekeeping 8. Funds Distribution	Testing 30. Achievement Test 31. Competency Test 32. Proficiency Test 42. Performance Test
H. S. Graduation Requirements **1. Credit Requirements** 9. Exit Test 10. Attendance	H. S. Graduation Requirements **1. Credit Requirements**
Teacher Standards Policies	Teacher Standards Policies
Entrance into Teacher Education 11. Test 12. GPA 13. Other	
Teacher Education Curriculum 14. Approved Program 15. Distribution Requirements	
Completion of Teacher Education 16. GPA 17. Basic Skills 18. Prof. Skills 19. Subject Specialty	
Entry-level Certification **2. Basic Skills Test** **3. Professional Skills Test** **4. Subject Specialty Test** 20. General Knowledge Test 21. Evaluation of Beginning Teacher 22. Approved Program	Entry-level Certification **2. Basic Skills Test** **3. Professional Skills Test** **4. Subject Specialty Test** 39. In-class Observation Elementary/Secondary Teacher Licensing 33 (36). Course Credits 34 (37). Teaching Methods in Math 35 (38). Supervised Teaching Experience
Recertification Requirements 23. Years of Experience 24. Formal Education 25. In-Service 26. Staff Development	40. Recertification Requirements 41. Advanced Professional Certificate 43. Teacher Certification Program for Persons from Non-education Field

The Rasch measurement model of state education policy-making specifies the probability of state n with activism b_n giving responses X_{ni} to policy i with difficulty d_i as

$$P\{X_{ni}\} = \frac{\exp[X_{ni}(b_n - d_i)]}{[1 + \exp(b_n - d_i)]}$$

where $X_{ni}=0$ when the policy is not enacted and $X_{ni}=1$ when the policy is enacted.

Then, the Rasch measure of state activism in education reform is estimated in a way that minimizes the difference between observed value (X_{ni}) and expected value ($P\{X_{ni}\}$) as follows:

$$\text{Logit} = \log[\frac{P\{X_{ni}=1\}}{P\{X_{ni}=0\}}] = (b_n - d_i) \text{ -> } \min \sum_n \sum_i (\mathbf{X}_{ni} - \mathbf{P}\{\mathbf{X}_{ni}\})^2$$

The logit is a "log odds" unit. Both state activism (b_n) and policy difficulty (d_i) are measured on the same logit scale. The difference between a state measure and policy difficulty is equal to the log odds of the state's probability of enacting the policy.

The four policies common to both test forms are used to equate the scale constructed from the 1991 data with the measures reported for 1984 (See policies 1-4 in Table 1). For the combined test with 43 policies, the state separation reliability is moderate (reliability = .75), and the policy separation reliability is high (reliability = .93).[1] In operationally defining state reform as an independent variable, one major concern is whether state educational policies as observed in the early 1980s have survived during the last decade so that the potential impact of reform policies on instructional practices as late as 1992 can be meaningfully examined. Input-oriented reform was expanded to include content-driven reform, and some states became more active than others. Thus, we need to

[1]The sample reliability of policy (item) separation is determined by the extent to which policy (item) calibrations are sufficiently spread out to define distinct levels along a variable. Only if items are clearly separated can we identify a direction along which measures can be interpreted.

differentiate between states that were high on both 1984 and 1991 reform measures and those that were low on both measures.

States that are commonly available for the 1990 and 1992 NAEP TSA data are selected and classified into three groups based on the 1984 and 1991 state policy measures: top quartile, middle half, and bottom quartile (See Table 2). For example, top-quartile states can be characterized by a relatively more active adoption of standards-based reform policies throughout the 1980s and early 1990s.

Table 2. Average Policy Measures of Three Groups of States

State grouping by level of reform	'84 Reform Activism	'91 Reform Activism
Top Quartile (Most Active)	.75	1.02
Middle Half	-.35	.64
Bottom Quartile (Least Active)	-1.50	-.35

Note. The scale for state policy measure is centered at zero logit.

Measuring Teachers' Progressive Instructional Practices as an Outcome Variable

The above-mentioned state reform policies are expected to affect instructional practices by uprgrading school curriculum and teacher quality as well as pushing students towards taking more advanced courses and demonstrating their academic proficiency. Student-centered instructional practices with a strong emphasis on higher-order thinking skills can be considered positive signs of implementation of many recent recommendations

for the reform of school mathematics. I extracted information on classroom activities from the NAEP TSA teacher questionnaire that was administered to eighth-grade math teachers. Teachers were selected if they taught the student the subject in which the student was assessed.[2] The following items from the 1990 and 1992 NAEP TSA teacher survey data are used to measure "progressive instruction" in an 8th grade math class:[3]

[1] How much emphasis on reasoning/analysis? (T031511/T044608)

[2] How much emphasis on communicating math ideas? (T031512/T044609)

[3] How often do students work in small groups? (T031403/T044503)

[4] How often do students write reports/do projects? (T031410/T044508)

[5] How often do students use measurement and geometry? (T031404/T044512)

[6] How often do students use calculators? (T031405/T044505)

[7] How often do students use computers? (T031406/T044506)

[8] How often do students write about problem-solving? (NA/T044507)

[9] How often do students discuss math with other students? (NA/T044509)

[10] How often do students work real-life math problems? (NA/T044510)

[11] How often do students make up math problems? (NA/T044511)

[12] How often assess students with written responses? (NA/T044703)

[13] How often assess students with projects/portfolios? (NA/T044704)

Rasch measurement model is also used to create a construct of "progressive instruction" from the above-mentioned survey data and to equate the two tests of different years and subjects. In order to examine instructional change over time at the state level,

---

[2] The purpose of drawing these samples was not to estimate the attributes of the teacher population, but to estimate the number of students whose teachers had various attributes and to correlate student characteristics and performance with the characteristics of their teachers (Johnson et al., 1994. The NAEP 1992 Technical Report, p. 86).

[3] Original variable names in the dataset appear in parenthesis: the items in the 1990 data precede their counterparts in the 1992 data. 'NA' indicates the absence of matching items in the dataset.

independent samples of teachers as linked to students in each state are tested two times, in 1990 when the NCTM standards were introduced, and in 1992 when the standards were expected to be much in place. There is also some corresponding change in the content of survey items on instructional practices: the 1992 test adds more NCTM-based practice items (e.g., problem-solving and application skills, and performance-based assessments) to the 1990 test. However, the two tests conduct parallel assessments to provide linkages between 1990 and 1992: there are seven common items covering instructional emphasis on reasoning and communication, use of technology, small-group work on projects (See items [1] through [7] above). BIGSTEPS, Rasch measurement program, is used to construct objective measures from the responses of 20,319 teachers over the 3- or 4-point scale items.

The seven items common to both test forms are used to equate the scale constructed from the 1992 data with the measures reported for 1990. The results of the co-calibration show a perfect item separation (reliability=1). In other words, items are very well separated in terms of the difficulty of practicing those instructional practices. On the other hand, teacher separation reliability is modest (reliabilty=.69). Since NAEP data are inappropriate for teacher-level analyses, teachers' measures of progressive instruction are matched to their students and aggregated to produce school-level and state-level average values.

Linking State Policy Measures to Classroom Instruction Measures

Did classroom instruction change from 1990 to 1992 as a result of state reform policies adopted during the 1980s and early 1990s? When the 1992 state average measure of progressive instruction is compared against the 1990 state average, 14 states appear to have advanced between the two years while 19 states retreated (See Figure 1). Interstate variation in the two-year instructional change seems to be somewhat associated with the

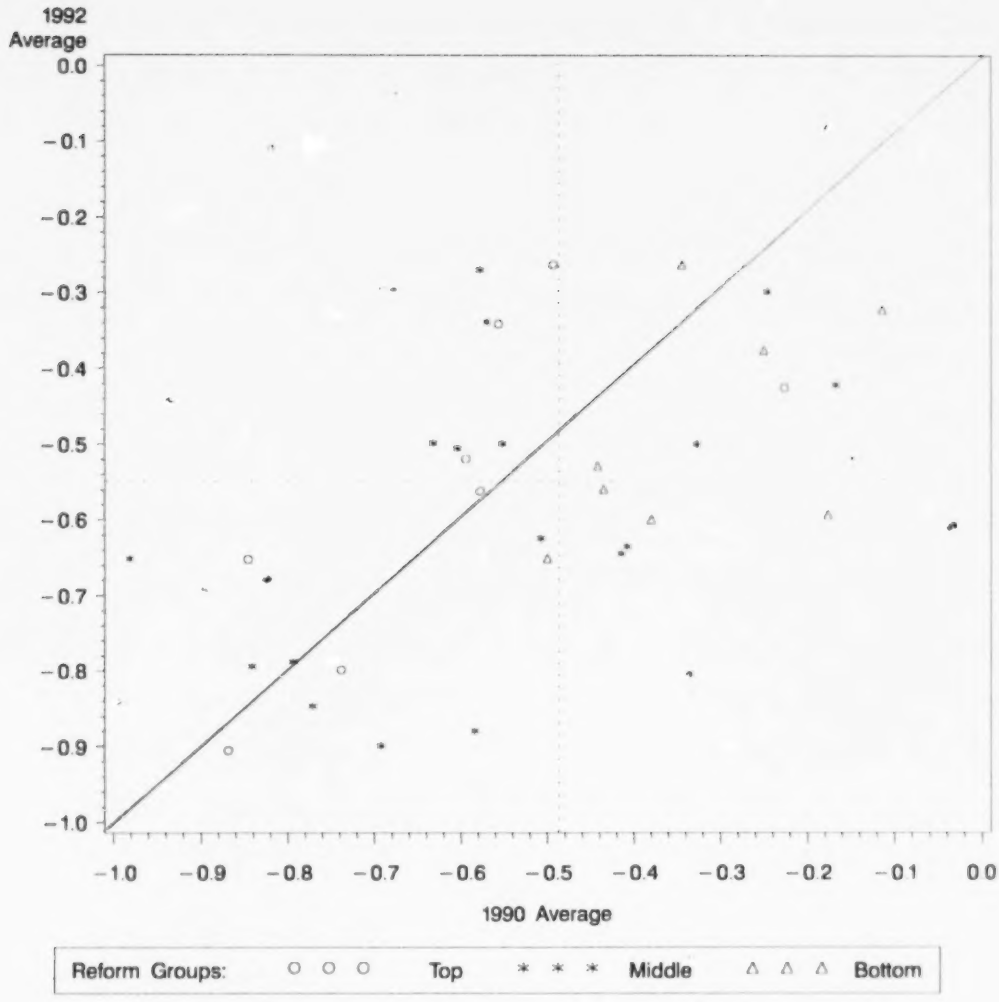

Figure 1. Plot of 1992 against 1990 state average measure of progressive instruction in mathematics, the mean of each year's measures indicated by broken lines

state's status of standards-based education reform. To see the relationship between state policies and instructional practices, states were classified into three reform groups according to their measures of reform activism. Indeed, when Figure 1 is evaluated in the context of error for each state, 10 of the 14 states (4 top and 6 middle) made statistically significant progress while 17 of the 19 states (2 top, 8 middle, and 7 bottom) showed statistically significant regression. The decline of more states between 1990 and 1992 in

progressive instruction may be attributed to change in the content of tests used: the 1992 test adds more challenging items to the 1990 test. Thus, variation among states in the extent of instructional change becomes of research concern regardless of the direction of change (positive vs. negative).

In examining the impact of state reform on instructional practices, we need to take into account within-state variation as well as between-state variation. Thus, my strategy is to conduct a multi-level analysis of the relations between policies and practices by capturing the relevant properties of school-level and state-level variables. First, using a sample of schools from each state (2,707 schools in 33 states), a school-level linear regression model is estimated for each school in each state to predict the association of school characteristics with progressive instructional practices as follows (See Appendix for a description of predictors):

Progressive instruction = f (Socioeconomic Status, Percent White, Professional Training, Teacher Autonomy, Ability Grouping, Academic Community, Program Activities, Absence of Problems, Urban Location, Rural Location)

Simultaneously, a state-level regression model is estimated for 33 states to predict the association of the perceived policy impact on actual instructional change. Instructional changes that principals attributed to content-driven policies are related to instructional practices reported by teachers. In order to control for past instructional practices at the state level, the 1990 state average measure of progressive instruction is included as a predictor. Some may question whether instructional change over two years can be meaningfully ascribed to policy effects. If the impact of state education reform on instructional practices had already occurred before 1990 and much of new instructional practices were in place by the end of last decade, the 1990 progressive instruction variable would be far from a "pure" pre-treatment measure that is a prerequisite for an appropriate adjustment variable in an

analysis of policy effects. Nevertheless, the validity of controlling for the 1990 status of instructional practices relies on the observation that most states did not attempt to substantially address the issues of curriculum and instruction until late 1980s or early 1990s. Specifically, I pose the following between-state model:

State mean progressive instruction = f (90 Math Instruction, Dummy for Middle Half, Dummy for Top Quartile)

As seen in Table 3, there is much greater variation among schools than among states (92.8 vs. 7.2). At the school level, organizational capacity for bottom-up change (Professional Training, Teacher Autonomy, and Program Activities) as well as social composition (Socioeconomic Status) are all positively related to progressive instruction, whereas schools that have high percent of whites and adopt ability grouping policy show less progressive instruction. Despite the positiveness of relationship, teaching and learning environment (Academic Community and Absence of School Problems) is not significantly related to the level of progressive instruction.

At the state level, the difference between top and bottom quartile states in progressive instruction turned out to be statistically insignificant (See Dummy for Top under Mean Outcome). It suggests that standards-based education reform may have failed to bring about substantial change in classroom practices at least during the early 1990s. Nevertheless, state reform turned out to make significant differences in the effects of some school-level variables on progressive instruction. The positive effect of professional development on instructional practices is stronger in top quartile states than in bottom quartile states (See Dummy for Top under Professional Training). This indicates that teacher certification and development policies may have been linked to state curricular/instructional standards. In addition, the instructional advantage of urban schools

Table 3. HLM Results: Final Analysis of 1992 Progressive Instruction in Math Class

Estimated Effects				
	Coefficients	Standard Error	t-Statistic	p-Value
State-level Effects				
Mean Outcome	-.577	.075	-7.672	.000
90 Math Instruction	.131	.039	3.355	.003
Dummy for Middle	.016	.091	.176	.862
Dummy for Top	.061	.105	.586	.562
School-level Effects				
Socioeconomic Status	.080	.017	4.656	.000
Percent White	-.062	.017	-3.582	.000
Professional Training	.126	.025	5.096	.000
Dummy for Middle	.045	.029	1.558	.130
Dummy for Top	.186	.034	5.429	.000
Teacher Autonomy	.055	.013	4.152	.000
Ability Grouping	-.100	.028	-3.548	.002
Academic Community	.028	.015	1.916	.065
Program Activities	.050	.015	3,370	.002
Absence of Problems	.022	.015	1.477	.150
Urban Location	-.175	.109	-1.609	.118
Dummy for Middle	.288	.116	2.473	.020
Dummy for Top	.388	.123	3.145	.004
Rural Location	.060	.033	1.815	.079

The Variance Table				
	Estimated Variance	Degrees of Freedom	Chi-Square	p-Value
state-level	.027	29	200.27	.000
school-level	.414			

	Percent variance partitioned by unconditional model	Percent variance explained by final model
state-level	7.2	28.9
school-level	92.8	15.1

over suburban counterparts is greater in top quartile states than in bottom ones (See Dummy for Top under Urban Location). This indicates that standards-based accountability may have made urban schools more aggressive in ensuring opportunity-to-learn.

Conclusion

Faced with needs for evaluating systemic school reforms in non-experimental settings, policy analysts and program evaluators are required to capitalize on variations that occur naturally in educational policy and practice. But the central question is how to measure and analyze such ambiguous and complex variations that result from the adoption and implementation of multiple policies for the entire school system. While objective measurement and multilevel analysis methods have been developed and found useful for educational research, they also have the potential to serve policy-oriented evaluation research. This study explores a comparative approach to evaluating systemic reforms through integrated application of the Rasch measurement and HLM analysis methods to existing state policy and classroom practice datasets.

The idea of comparng two groups of states on their policy outcome measures is similar to nonequivalent control group design (see Campbell and Stanley, 1963) in that the most active reform states can be regarded as experimental group, and the least active states as comparison group. Nevertheless, the research design proposed in this paper differs from the nonequivalent control design in some significant ways: 1) treatment is not a single, independent program but a set of interrelated programs, 2) group exposure to given treatment is not simply a question of all versus nothing but rather matter of degree, 3) all of the programs that constitute treatment do not have to occur between pre-test and post-test, but some of them may begin before pre-test and continue through post-test, and 4) subjects that are examined on pretest and postest do not have to be the same but instead they can be

sampled independently. Thus, the propopsed evaluation approach should give more flexibilities for evaluation design in real-life settings but at the same time more difficulties for interpretation of evaluation results.

The illustrated study of state policy evaluation raises some substantive and methodological concerns to be addressed both at the measurement and analysis stages. On the measurement front, the study relies on survey data to construct measures of state policies and classroom practices. But the survey instruments used have some limitations. The ETS and CCSSO state policy surveys focused on the type or level of state policy activities but could not capture variation in the content or function of adopted policy instruments. Likewise, the NAEP school teacher survey tells us much about the frequency or intensity of certain instructional practices but nothing about the quality or meaning of those practices for students. To cope with those problems, it is necessary to conduct more sophisticated policy/practice survey and complement large-scale survey-based data analyses with in-depth case studies.

On the analysis front, the policy evaluation study focuses on interstate comparisons. But even within a single state, there are significant possibilities for comparative study. Many state policies or programs are carried out through a series of local projects, with local variations in strategy and procedure. Cross-program study—that is, evaluation of all or a sample of the local projects—can yield information on the relative success of different methods of program implementation for the attainment of the common goals. Thus, it is useful to see how schools in a state that adopts systemic reforms vary in translating state curriculum and assessment policies into their own programs to improve instructional practices and student outcomes.

References

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models. Newbury Park: Sage Publication.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Lee, J. (1996). Multilevel linkages of state education reform to instructional practices. Paper presented at the annual meeting of American Educational Research Association.

Lee, J. (1997a). State activism in education reform: Applying the Rasch model to measure trends and examine policy coherence. Educational Evaluation and Policy Analysis 19 (1), 29-43.

Lee, J. (1997b). Multilevel linkages between state policies and educational outcomes: An evaluation of standards-based education reform in the United States. unpublished doctoral dissertation. Chicago: The University of Chicago Press.

Slavin, R. E., Karweit, N.L., & Madden, N.A. (1989). Effective programs for students at risk. Boston: Allyn and Bacon.

Weiss, C. H. (1972). Evaluation research: Methods for evaluating program effectiveness. Englewood Cliffs, N.J.: Prentice-Hall

Wright, B. D., & Stone, M. H. (1979). Best Test Design: Rasch Measurement. University of Chicago: MESA Press.

Appendix. School-Level Predictors of Progressive Instruction

The following variables are constructed from the 1992 NAEP 8th grade mathematics teacher and school survey data. Each principal's or teacher's responses to 2 to 4-point scale items are transformed through principal component analyses into factor scores.

Absence of Problems: A factor composite of principals' reports about absence of schoolwide problems in the following aspects : student tardiness, absenteeism, cutting classes, physical conflicts, drug/alcohol, teacher absenteeism, racial and cultural conflicts, and student health (factor made from C032401-8). School-level factor loadings are as follows: C032401, .71; C032402, .71; C032403, .70; C032404, .74; C032405, .49; C032406, .62; C032407, .61; C032408, .64. Factor has an eigenvalue of 3.47 and explains 43 percent of the combined variance.

Communal Climate: A factor composite of teachers' reports about positiveness of school climate in the following aspects: teachers' relations with administration, teacher morale, student attitudes to academics, teacher attitudes to academics, parent support for academics, regard for school property, and relations between teachers and students (school-level average of factor made from C032501-7). Student-level factor loadings are as follows: C032501, .63; C032502, .71; C032503, .74; C032504, .69; C032505, .69; C032506, .67; C032507, .75. Factor has an eigenvalue of 3.40 and explains 49 percent of the combined variance.

Program Activities: A factor composite of principals' reports about school improvement activities in the following aspects: involving parents as aides in class, encouraging parents to visit classes, having minimum requirement for homework, performance-based competition system for teacher, mentoring program for teachers, before/after school remediation program, summer-school program, and dropout prevention program (factor made from C032207-8, C032301, C032303-6, C032314). School-level factor loadings are as follows: C032207, .49; C032408, .56; C032301, .33; C032303, .28; C032304, .45; C032305, .57; C032306, .54; C032314, .50. Factor has an eigenvalue of 1.81 and explains 23 percent of the combined variance.

Professional Training: A factor composite of teachers' reports about their training in the following areas: estimation, math problem-solving, use of manipulatives, use of calculators, students' math thinking (school-level average of factor made from T041701-2, T041708, T041704-5). Student-level factor loadings are as follows: T041701, .70; T041702, .67; T041708, .69; T041704, .68; T041705, .65. Factor has an eigenvalue of 2.31 and explains 46 percent of the combined variance.